Improvements in or Relating to Screening for Papilloma Viruses

Field of the Invention

This invention relates to a method of screening for precursor lesions which can lead to cervical malignancy, methods of detecting and typing HPV infections, and reagents of use in the above methods.

Background of the Invention

Papillomaviruses (PVs) cause epithelial tumours in humans which vary in severity depending on the site of infection and the HPV (human papilloma virus) type involved (Laimins-, 1993; Villiers de, 1994). Low risk types such as HPV 1 or HPV63 (Egawa *et al*, 1993a; Egawa *et al*, 1993b) cause benign cutaneous warts which progress to malignancy only rarely, while high risk viruses such as HPV 16 and HPV31 cause flat warts at mucosal sites, and are associated with high grade cervical intraepithelial neoplasia (CIN) and cancer (Schneider, 1994). Formation of an HPV-induced tumour is thought to require infection of an epithelial basal cell, and the expression of viral early proteins in order to stimulate cell proliferation. The late stages of the virus life cycle, which ultimately lead to the production of infectious virions, are initiated only as the infected cell migrates through the upper differentiated layers of the epidermis. Viral and cellular events which influence HPV late gene expression have not been characterised as, until recently, there has been no convenient system for mimicking productive infection *in vitro* (Laimins, 1993)

Studies on naturally-occurring warts have revealed the virus to encode three late proteins - L1 and L2, which are virion coat proteins (Doorbar *et al*, 1987), and E1^E4, a non-structural late protein of unknown function (Doorbar et al, 1986). In HPV1-induced warts the E1^E4 protein is first expressed in cells of the lower spinous layer, and assembles into distinctive cytoplasmic and nuclear inclusions. During terminal differentiation it is post-transcriptionally modified by phosphorylation (Grand *et al*, 1989) and by removal of sequences from the N-terminus (Doorbar *et al*, 1988; Roberts *et al*, 1994). The E1^E4 proteins of high risk viruses have been poorly characterised,

because it has been thought that HPV16-induced lesions contain only small numbers of productively infected cells, and that these contain only low levels of E4 (Doorbar *et al*, 1996b: Crum *et al*, 1990). A single Mab (TVG 402) to HPV16 EI^E4 has been used to locate the protein to the cytoplasm but was reported not to work well on paraffin-embedded archival material (Doorbar *et al*, 1992). Furthermore, polyclonal antibody studies on the E4 proteins of mucosal viruses have yielded conflicting results. One study has supported the above findings (Crum *et al*, 1990), while another has indicated that the protein is located to the nucleus (Palefsky *et al*. 1991).

In many countries there are screening programmes to detect the presence of cervical carcinoma at an early stage. Generally such programmes operate by obtaining cervical smears from women potentially at risk of developing cervical cancer, with the resulting smears routinely being examined by conventional histopathological techniques. These techniques are laborious and time-consuming, require considerable experience to interpret results correctly, and frequently give rise to relatively large percentages of false positive results, causing unnecessary alarm. False negatives can occur when screening is carried out by inexperienced personnel and can lead to the classification of pre-cancerous lesions as normal. There is thus a need for an improved cervical cancer screening method.

It is well known that there is a very strong correlation between HPV-infection and development of cervical carcinoma: over 90% of women with cervical carcinoma show evidence of HPV infections of the cervix. Accordingly, one possible alternative to conventional histopathological examination of cervical smears is to examine samples for evidence of HPV infection. For example, there have been numerous proposals to screen for cervical carcinoma by performing DNA hybridisation assays on samples, using nucleic acid probes specific for HPV sequences. Such hybridisation assays are generally favoured by those skilled in the art, because of the ready availability of suitable reagents and because of their high specificity.

Thus, for example in Fields Virology (Fields *et al*, [Eds.] *Virology* Vol. 2, p2099, 3rd Edn. (1996) Raven Press, New York), an authoritative virology text book, it is stated that "Diagnosis of an HPV type in a tissue requires nucleic acid hybridization studies".

In contrast, screening for cervical carcinoma by detection of expression of HPV polypeptides has generally been disregarded, being considered unsuitable for a number of reasons, primarily because of the difficulty in obtaining suitable reagents and, more significantly, many HPVs produce very little virus protein in mucosal infections, making detection difficult, uncertain and unreliable. Thus, in Fields Virology (cited above) it is stated that "immunologic detection of viral capsid antigens" is "of limited value". The possibility of immunologic detection of other viral antigens is not even considered. If one were to develop a screening method based on detection of expression of viral proteins, the most likely choice of target would be those proteins which are best-characterised, such as L1 or L2. The function of E4 protein is at present unknown. Its expression pattern in cervical lesions has not been determined conclusively in the prior art so the molecule has not been an obvious choice for selection as a target for detecting HPV infection.

Summary of the Invention

In accordance with the present invention, it has now been demonstrated that HPV infection can be detected in a sample taken from a patient by using molecules which bind specifically to E4 protein of HPVs. In particular, the invention provides a method of screening samples for pre-cancerous cervical lesions, using molecules which bind specifically to HPV E4 protein.

The present studies have clearly demonstrated HPV16 E4 protein to be cytoplasmic, and to be produced in cells supporting vegetative viral DNA replication.

In a first aspect the invention provides a method of detecting a papilloma virus infection in an organism, the method comprising the steps of: obtaining a sample of the

organism's cells from the site of potential papilloma virus infection; contacting the cells with a molecule that binds specifically to papilloma virus E4 protein; and monitoring said binding.

In particular, the invention provides a method of screening for pre-cancerous cervical lesions, comprising the steps of: obtaining a sample of cervical cells from a subject; contacting the cells with a molecule that binds specifically to HPV E4 protein; and monitoring said binding.

Moreover, the invention provides a method of determining the type(s) of HPV infection in a patient, the method comprising the steps of: obtaining a sample of the patient's cells from the site of HPV infection; contacting the cells with a molecule that binds specifically to a subset of HPV E4 proteins; and monitoring said binding.

In a further aspect the invention provides an antibody molecule, or an antigen-binding variant thereof, which binds specifically to HPV E4 protein in the region of amino acid residues RPIPKPSPWAPKKHRRLSSDQDSQTP of HPV16 E4 protein, or the corresponding hydrophilic, acid/base-rich region of other HPV E4 proteins.

The invention moreover concerns the use of molecules capable of binding to E4 to target antiviral agents capable of destroying papilloma viruses and/or cells infected by papilloma viruses. Such molecules may be antibodies or peptides as described above and exemplified herein, optionally conjugated to anticancer or antiviral agents.

Brief Description of the Drawings

Figure 1A shows the amino acid sequence of HPV16 E4 protein and the binding sites of various antibody molecules or E4-specific antigen-binding fragments of antibodies;

Figure 1B shows the sequence of the E4 protein from HPV16 (top row), HPV1 (bottom row) and a consensus sequence (middle row), and the binding sites of various antibodies or antigen-binding variants of antibodies;

Figures 2A-2D show four sensograms (arbitrary response units against time in seconds) obtained using surface plasmon resonance apparatus;

Figures 3-8 are micrographs showing variously stained samples, as explained in the text; and

Figure 9 is an amino acid sequence alignment of part of HPV E4 proteins.

Detailed Description of the Invention

The method according to the present invention permits the detection, identification and diagnosis of papilloma viruses and papilloma virus infections in organisms susceptible to such infections.

Such organisms are preferably mammals, and most preferably humans. Where the organism is a human organism, the papilloma virus may be a type or types of human papilloma virus (HPV).

The sample of patient's cells may comprise skin cells (e.g. in the case of warts, veruccas and the like, caused by cutaneous HPV infections). Cutaneous lesions, such as those induced by HPV types 5, 8, 14, 17, 20, are difficult to manage clinically, and are often associated with malignancies in immunosuppressed patients (Benton *et al*, 1992 Papillomavirus Reports 3, 23-26). Alternatively, the sample may comprise mucosal cells, especially cervical cells, in the case of HPV infections of the urinogenital tract. Methods of obtaining and preparing such samples for use in the

method of the invention are known to those skilled in the art or will be apparent from the present disclosure.

The term "pre-cancerous cervical lesions" is intended to refer to those abnormalities which clinically may be described as "pre-malignant" conditions and which may, without treatment, proceed to full malignancies. As set forth above, such lesions are screened for routinely by, for example, cervical smear testing. The present invention allows for cells obtained from patients by methods such as cervical smears to be tested more accurately and more quickly for HPV infection.

Preferably, the molecule which binds specifically to E4 protein comprises an antibody molecule or an antigen-binding variant thereof, such as an Fab, Fv, scFv, "diabody" and the like. The molecule may comprise monoclonal or polyclonal antibodies, or antigen-binding portions of antibodies selected from libraries by screening (e.g. using phage display technology). Alternatively the molecule may be some other polypeptide, peptide, a synthetic compound or an RNA or DNA aptamer, generated by a procedure such as SELEX. In some preferred embodiments the molecule comprises a label moiety, such as a fluorophore, chromophore, enzyme or radio-label, so as to facilitate monitoring of binding of the molecule to E4 protein. Such labels are well-known to those skilled in the art and include, for example, fluorescein isothiocyanate (FITC), β-galactosidase, horseradish peroxidase, streptavidin, biotin, ^{35}S or ^{125}I. Other examples will be apparent to those skilled in the art. The label may in some instances be conjugated to the antibody or antigen-binding variant, or may be present (where the label is a peptide or polypeptide) as a fusion protein.

Preferably the molecules used in the method of the invention bind selectively to the E4 protein of a certain HPV type or types, but not to the E4 protein of other HPV types. Accordingly, in one embodiment the invention can be used to determine the type or types of HPV infecting a patient. This is very significant, as progression to malignant disease (and hence clinical prognosis) is heavily dependent on HPV type. Accordingly, in a second aspect the invention provides a method of determining the type(s) of HPV

infection in a patient, the method comprising the steps of: obtaining a sample of the patient's cells from the site of HPV infection: contacting the cells with a molecule that binds specifically to a subject of HPV E4 proteins; and monitoring said binding.

In the method of the second aspect of the invention, the subset of E4 proteins to which the molecule binds may consist of a single HPV type E4 protein, or may consist of a plurality of E4 proteins of different types, but will not encompass the E4 proteins of all known HPV types, such that binding or non-binding (as appropriate) of the molecule to the E4 protein present in the cell sample will allow an investigator to make certain deductions about the identity of the HPV type(s) infecting the patient.

In practice it may be advantageous to employ a plurality of different molecules, which bind to different subsets of E4 proteins. This may be necessary to identify unambiguously the type(s) of HPV infecting the patient, although it may not be essential as a prognostic indicator. For example, the ability to limit the infecting HPV type(s) to a particular subset (or exclude such a subset) may be sufficient. By way of explanation, it is known that mucosal HPV types 6, 11, 42, 43 and 44 are associated with external genital papillomas (condylomata accuminata) which have a low risk of progression to cancer, but are difficult to eradicate and are disruptive to the lives of the patients. The higher risk mucosal types (31, 33, 35, 51, 52, 58, 61 and 16, 18, 45, 56) cause asymptomatic flat warts (flat concyloma) which can progress to high grade cervical intraepithelial neoplasia (CIN) and cancer. The highest risk of progression to malignancy is associated with lesions caused by HPV types 16, 18, 45 and 56.

Molecules which bind to desired HPV types, but not to undersired HPV types, may be generated for example by randomisation and selection techniques. These include phage display, and other techniques suitable for displaying antibodies or other polypeptides; and procedures for generating nucleic acid binding molecules, for example RNA aptamers, such as SELEX. These procedures are well known to those of ordinary skill in the art and described below for the purposes of exemplification. The invention

accordingly provides HPV-binding molecules targetted to the HPV E4 protein, which are useful in methods as described herein.

According to the present invention, E4-binding molecules are preferably targeted to extracellular portions of the E4 polypeptide. Such portions tend to be hydrophilic in character. Preferably, therefore, the E4 binding molecules according to the invention specifically bind to hydrophilic portions of the HPV E4 protein.

The present invention moreover provides a particular region of the E4 protein to which molecules (particularly antibody molecules or variants thereof) may bind with considerable specificity. Although homologous regions exist in all HPV E4 proteins, the region varies in amino acid sequence between HPVs of different types. The region corresponds to a peak of hydrophilicity in the E4 protein and is probably surface-exposed. The region is highly charged (acid/base-rich). In HPV type 16, the amino acid sequence of the region is (from N-terminal to C-terminal) RPIPKPSPWAPKKHRRLSSDQDSQTP. Clearly the amino acid sequence of the E4 proteins of other HPV types will not necessarily be identical to that in type 16, but with the benefit of the present disclosure (e.g. figure 9) the corresponding region can readily be identified in other E4 proteins by those skilled in the art by use of conventional alignment and sequence comparison computer programs (about 65 of the 70 or so known HPV genomes have been cloned and sequenced).

Thus, in a third aspect the invention provides an antibody molecule, or an antigen-binding variant thereof, which binds specifically to HPV E4 protein in the region of amino acid residues RPIPKPSPWAPKKHRRLSSDQDSQTP of HPV16 E4 protein, or the corresponding hydrophilic, acid/base-rich region of other HPV E4 proteins, preferably other than the antibody TVG 402 identified by Doorbar *et al*, (1992 Virology 187, 353-359).

Moreover, the invention provides the use of an antibody molecule, or an antigen-binding variant thereof, which binds specifically to HPV E4 protein in the region of

amino acid residues RPIPKPSPWAPKKHRRLSSDQDSQTP of HPV16 E4 protein, or the corresponding hydrophilic, acid/base-rich region of other HPV E4 proteins for the detection of HPV infections as described herein.

The corresponding hydrophilic acid/base-rich regions of large numbers of different HPV types are shown in Figure 9. Figure 9 shows a consensus-type amino acid sequence ("most likely") on the top row, with the sequence of HPV E4 proteins below. Dots indicate gaps introduced to facilitate the alignment, dashes denote amino acid residue matches with the consensus sequence. Numbering on the right hand side of the figure indicates the number of amino acid residues from the actual or predicted E1^E4 splice site. It will be appreciated by those skilled in the art from the alignment that whilst the hydrophilicity of the region is conserved amongst different HPV types, the actual amino acid sequence varies quite considerably, such that reagents binding to this region may be expected to be highly HPV type-specific.

Preferably the antibody of the invention has a binding site, as identified by the SPOTS epitope mapping system, within the region RRIPKPSPWAPKKHR (or the corresponding amino acid sequence from other HPV types). A particularly preferred molecule is the Fab fragment TVG405, describ
epitope PKPSPWAPKKH(R) with extremely high
in the methods of the invention defined above.

SEQ ID NO 167

SEQ ID No: 168

The arginine residue indicated in brackets at the
not essential for high affinity binding.

The Fab fragment TVG405 was isolated by th
technology, as described below. Those skilled
antibodies or Fab fragments may readily be c
techniques (and screening with E4 proteins or portions thereof), or by using more conventional immunisation techniques (e.g. immunising mice, rabbits, rats or the like with E4 protein or peptides corresponding to portions of the E4 protein) to obtain

polyclonal antisera or monoclonal antibodies (using well known hybridoma techniques of Milstein et al). Complete antibody molecules can readily be prepared from Fab - encoding sequences (e.g. isolated by phage display techniques) using standard DNA manipulation techniques described by Sambrook *et al*, (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, NY, USA) to join appropriate DNA sequences.

Similarly, standard DNA manipulative techniques can be used to modify DNA sequences encoding anti-E4 antibodies or antigen-binding variants thereof. In particular site-directed mutagenesis or PCR can be used to modify the coding sequences, so as to produce modified anti-E4 antibodies with different binding specificities or affinities. Alternatively fusion proteins. comprising the E4-binding site of an Fab, Fv or antibody and the like, may be prepared.

Molecules capable of binding E4 may be used as anti-viral or anti-cancer agents, or parts of such agents. For example, antibody molecules or E4-binding peptide as described above may be employed for this purpose. Preferably, however, the E4 protein and/or molecules capable of binding thereto may be used to design E4-binding molecules, preferably small molecules, by rational drug design.

Such a process preferably involves the crystallisation of E4 or a molecule capable of binding thereto. More preferably, such a process involves the co-crystallisation of E4 and a binding agent. Such a procedure gives information concerning the interaction between E4 and the binding molecule, which can be used to design small molecules capable of mimicking the binding interaction.

Crystallisation involves the preparation of a crystallisation buffer, for example by mixing a solution of the peptide or peptide complex with a "reservoir buffer", preferably in a 1:1 ratio, with a lower concentration of the precipitating agent necessary for crystal formation. For crystal formation, the concentration of the precipitating agent is increased, for example by addition of precipitating agent, for example by titration,

or by allowing the concentration of precipitating agent to balance by diffusion between the crystallisation buffer and a reservoir buffer. Under suitable conditions such diffusion of precipitating agent occurs along the gradient of precipitating agent, for example from the reservoir buffer having a higher concentration of precipitating agent into the crystallisation buffer having a lower concentration of precipitating agent. Diffusion may be achieved for example by vapour diffusion techniques allowing diffusion in the common gas phase. Known techniques are, for example, vapour diffusion methods, such as the "hanging drop" or the "sitting drop" method. In the vapour diffusion method a drop of crystallisation buffer containing the protein is hanging above or sitting beside a much larger pool of reservoir buffer. Alternatively, the balancing of the precipitating agent can be achieved through a semipermeable membrane that separates the crystallisation buffer from the reservoir buffer and prevents dilution of the protein into the reservoir buffer.

In the crystallisation buffer the peptide or peptide/binding partner complex preferably has a concentration of up to 30 mg/ml, preferably from about 2 mg/ml to about 4 mg/ml.

Formation of crystals can be achieved under various conditions which are essentially determined by the following parameters: pH, presence of salts and additives, precipitating agent, protein concentration and temperature. The pH may range from about 4.0 to 9.0. The concentration and type of buffer is rather unimportant, and therefore variable, e.g. in dependence with the desired pH. Suitable buffer systems include phosphate, acetate, citrate, Tris, MES and HEPES buffers. Useful salts and additives include e.g. chlorides, sulphates and further salts specified in Example 1. The buffer contains a precipitating agent selected from the group consisting of a water miscible organic solvent, preferably polyethylene glycol having a molecular weight of between 100 and 20000, preferentially between 4000 and 10000, or a suitable salt, such as a sulphates, particularly ammonium sulphate, a chloride, a citrate or a tartrate.

A crystal of E4 itself or an E4-derived peptide, or E4 (peptide)/binding partner complex according to the invention may be chemically modified, e.g. by heavy atom derivatization. Briefly, such derivatization is achievable by soaking a crystal in a solution containing heavy metal atom salts, or a organometallic compounds, e.g. lead chloride, gold thiomalate, thimerosal or uranyl acetate, which is capable of diffusing through the crystal and binding to the surface of the protein. The location(s) of the bound heavy metal atom(s) can be determined by X-ray diffraction analysis of the soaked crystal, which information may be used e.g. to construct a three-dimensional model of the peptide.

A three-dimensional model is obtainable, for example, from a heavy atom derivative of a crystal and/or from all or part of the structural data provided by the crystallisation. Preferably building of such model involves homology modelling and/or molecular replacement.

The preliminary homology model can be created by a combination of sequence alignment with any of the E4 proteins the sequence of which is known, secondary structure prediction and screening of structural libraries. For example, the sequences of HSV 16 and 34 E4 can be aligned as set forth herein.

Computational software may also be used to predict the secondary structure of E4 peptides or peptide complexes. The peptide sequence may be incorporated into the E4 structure. Structural incoherences, e.g. structural fragments around insertions/deletions can be modelled by screening a structural library for peptides of the desired length and with a suitable conformation. For prediction of the side chain conformation, a side chain rotamer library may be employed.

The final homology model is used to solve the crystal structure of E4 or peptides thereof by molecular replacement using suitable computer software. The homology model is positioned according to the results of molecular replacement, and subjected to

further refinement comprising molecular dynamics calculations and modelling of the inhibitor used for crystallisation into the electron density.

Similar approaches may be used to crystallise and determine the structure of E4-binding polypeptides, including antibodies and antibody fragments, for example those provided by the present invention.

It has surprisingly been found that E4 expression correlates strongly with vegetative DNA replication in HPV-infected cells, making detection of E4 expression a particularly appropriate indicator of HPV infection, and thus particularly useful in screening for precancerous cervical lesions.

Present available methods of cervical screening by HPV detection are based on DNA hybridisation. They involve cell lysis or permeabilisation and are performed in an ELISA-type 96 well format. The hybridisation is ultimately visualised as a colour change in one of the wells.

Although the antibodies of the present invention could be used in a similar way (i.e. following cell lysis), they are amenable to a quicker procedure which would be more readily carried out routinely by histopathology laboratories. Samples comprising cervical cells may be taken as usual. These are be spread for example on a microscope slide or other support using techniques known in the art, for example as exemplified herein, and stained with, for example, an anti-E4 Fab. Detection may be performed with a secondary antibody-enzyme conjugate (horseradish peroxidase, alkaline phosphatase), or the Fab could be directly conjugated, for example to a fluorophore, such as FITC. This approach may be adapted for use with systems that are currently available for increasing the sensitivity of antibody detection. At present, cervical smears are examined routinely by microscopy. The proposed approach would require no new equipment and could easily fit around existing methods.

It is envisaged that the standard method of detection may be modified. Antibody binding may be carried out while the cells are in suspension, with cells being spun down prior to analysis. This would improve the quality of the screen.

Considerable effort in diagnosis is aimed at automating screening methods. The use of antibodies or antigen-binding variants thereof for HPV detection greatly facilitates this.

In summary, it has been shown that:

1. The E4 protein can be detected in productively infected HPV-induced lesions, and in low and high grade cervical neoplasia even when differentiation of the infected keratinocyte is insufficient to support production of capsid proteins and assembly of infections virions.

2. E4 expression correlates closely with vegetative viral DNA replication indicating that detection of the E4 protein is as efficient as detection of viral DNA replication for the detection of virus infection.

3. The E4 protein is abundant in the upper layers of infected tissue and is thus detectable in cells taken during routine smear tests.

The invention will now be described by way of illustrative examples .

Example 1

Preparation of Anti-E4 monoclonal and polyclonal immunoglobulins

Although Mabs against HPV16 E1^E4 have been described previously (TVG401, 402, 403; Doorbar *et al*, 1992) these reagents recognise a single overlapping epitope at the major antigenic site of E4, and have been reported not to detect the protein in archival tissue biopsies (Doorbar et al, 1992).

Although these results suggest that E4 may not be a candidate for immunological detection of HPV, further antibodies are generated targeted at the N and C termini of HPV16 E4.

The generation of further Mabs by standard hybridoma technology results in the isolation of TVG404, an IgM which recognises an epitope at the very C-terminus of the protein.

To complement this reagent polyclonal antiserum to the N-terminus of the protein is raised against an N-terminal synthetic peptide (-E4 N term). Polyclonal antibodies (to HPV16 and HPV63 E4 proteins) are prepared by immunisation of rabbits with maltose binding protein E4 fusion protein (MBP-E4). Antibody titres are monitored in ELISA using purified glutathione S transferase E4 fusion protein (GST-E4).

Antibodies to the N-terminus of the protein are raised against the synthetic peptide MADPAAATKYPLC after conjugation to thyroglobulin or keyhole limpet haemocyanin through its C-terminal cysteine residue. Conjugation is carried out using m-Maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) as described by Green et al (1982).

Antibody specificities are confirmed by epitope mapping, as follows: the HPV 16 E4 protein is synthesised as a series of 85 overlapping octamers (single amino acid overlap) by solid phase fmoc chemistry using the SPOTS epitope mapping system (Genosys Biotechnologies, Cambridge, UK). Accuracy of synthesis is confirmed using the HPV16 E1^E4 monoclonal TVG402 which binds the major antigenic site of the protein (Doorbar *et al*, 1992). Filters are regenerated as described by the manufacturers and antibody binding is visualised by ECL (Amershām, Little Chalfont, UK). Polyclonal serum is used at 1/250 dilution, purified Fabs at approximately 1 g/ml, and hybridoma supernatant at 1/10 dilution.

In Figure 1A the sequences of the 85 overlapping E4 synthetic peptides are shown at the top of the figure, and the results of the epitope mapping experiments are shown below. The dark spots represent binding of the antibody to the synthetic peptide shown above it. Only the portion of the filter containing peptides which react with each antibody are shown.

In Figure 1B, the locations of epitopes on the E1^E4 amino acid sequence are summarised above the HPV16 sequence. Alignment with a consensus E4 sequence prepared by comparison of 70 putative E1^E4 sequences (Doorbar and Myers, 1996b) is shown beneath the sequence of HPV16 E1^E4, and the sequence of the HPV1 E1^E4 protein is shown beneath this. The binding sites of the existing HPV1 E1^E4 Mabs (Doorbar *et al*, 1988) are shown beneath the HPV1 sequence. The proteolytic cleavage sites that give rise to the 16K and 10/11K gene products in the E1^E4 protein of HPV1 (Doorbar *et al*, 1988; Roberts *et al*, 1994) are indicated beneath the HPV1 sequence allowing prediction of putative sites in the E1^E4 sequence of HPV16.

Example 2

Preparation of Synthetic Immunoglobulins

Fabs are isolated from a synthetic antibody displayed on fd bacteriophage (Griffiths et al, 1994) as described below. Immunotubes (Life Technologies, Paisley, UK) are coated overnight at 4°C with either MBP-E4 or GST-E4 at a concentration of 0.1 g/ml. These are subsequently blocked at 37°C for 1 hour in PBS/2% marvel™ prior to incubation in the presence of 10^{11} phage on a blood tube rotator (37°C). Unbound phage are poured off and tubes are washed 10x with PBS/0.1% Tween 20. Binders are eluted with 100mM triethylamine pH 11.0 (1ml) and immediately neutralised with 1M Tris (pH8.0) before being reintroduced into *E. coli* TG1 cells. The enriched library is grown up and the selection procedure repeated three more times.

Phage selections are carried out alternately against GST 16 E1^E4 and MBP 16 E1^E4 in order to prevent isolation of antibodies to MBP or GST protein, using a repertoire of

6.5 x 10^{10} (Griffiths *et al*, 1994). MBP 16 E4 is expressed at higher levels (>50mg/litre of bacteria) than the GST fusion (approx. 5mg/litre of bacteria) but, in any event, antibody isolation requires as little as 1 g of antigen (Hawkins *et al*, 1992). Phage displaying antibodies with affinity for E4 are identified by ELISA (against GST-E4, MBP-E4, GST and MBP), and activity is confirmed by phage western blotting. Immunoglobulin genes are transferred from the isolated phage into the bacterial expression vector pUC119.His.myc (Griffiths *et al* 1994) and soluble Fabs are purified from the periplasmic space of induced bacteria by Nickel-NTA chromatography (Qiagen, Crawley UK). Antibody titres are established by ELISA.

After four rounds of selection, individual clones are examined for their ability to bind either E1^E4, unfused GST or MBP, or bovine serum albumin (BSA). 47 clones (out of 100) are able to bind MBP 16 E1^E4, of which 39 could also bind GST 16 E4. None of these clones interacted with BSA, GST or MBP. *Bst*NI fingerprinting (Marks *et al*, 1992: Nissim *et al*, 1994) revealed three distinct Fabs amongst these clones, and their immunoglobulin genes are subcloned into the prokaryotic expression vector pUC119His.6myc to allow the production of soluble anti-E4 Fabs (Griffiths *et al*, 1994). Approximately 5mg (per litre of bacteria) of anti-E4 Fab (TVG 405, 406 and 407) can be extracted from the periplasmic space of induced bacteria and all are found to specifically detect E1^E4 by ELISA and western blotting. Fab TVG 407 binds an epitope which is identical to that recognised by the hybridoma-derived Mab, TVG 409 (Fig 1). The remaining synthetic Fabs recognise novel epitopes upstream (TVG 405) or downstream (TVG 407) of this major antigenic region of E4 and the results are summarised in Figure 1.

It is found that the amino acid sequence of the CDR3 loops of the TVG 405 and TVG 407 Fabs are as follows:

TVG 405	heavy chain CDR3 sequence:	LLRGAFDY
	light chain CDR3 sequence:	NSRDSSGGNAV

TVG 407 heavy chain CDR3 sequence: LVQGSFDY
light chain CDR3 sequence: QADSSTHV

Measurement of Antibody Affinity

Affinities of synthetic (TVG405, TVG406 and TVG407), and hybridoma-derived Fabs (TVG402) are analysed by surface plasmon resonance using a BIAcore 2000 instrument (Pharmacia Biosensor, St. Albans, UK) as described by the manufacturer. MBP-E4 aggregates are dissociated under reducing conditions in 0.5% SDS, 1mM β-mercaptoethanol, 50mM $Na_2CO_3/NaHCO_3$ (pH 8.5) and biotinylated using NHS-LC-biotin (Sigma, St Louis, USA; 25mg/ml in DMSO) at a biotin:protein molar ratio of 6:1 (Johnson *et al*, 1991). Monomeric MBP-E4 is recovered by FPLC chromatography using a Superdex S200 HR10/30 column run in 6M Urea/1mM β-mercaptoethanol/PBS/0.2mM EDTA (pH7 2), before being bound to a streptavidin-coated sensor chip and "refolded" *in vitro* in PBS/0.2mM EDTA/0.1mg/ml protease-free BSA (Sigma). Fabs are isolated from purified TVG402 using an Immunopure Fab kit (Pierce, Rockford, USA), and monomeric preparations are obtained by FPLC gel chromatography (Superdex S200 HR10/30 column run in PBS/0.2mM EDTA (pH 7.2)) Sensor chip surfaces are regenerated using 6M urea column buffer (described above). On and off rates are derived by non linear curve fitting using the proprietary 'BIAanalysis' software.

Binding activities are in the order of 20% of total protein levels for the bacterially-derived antibodies, and 50% for Fabs derived from hybridoma culture supernatant. The affinities of TVG405 and TVG402 are calculated from on- and off-rates obtained by non-linear curve fitting to sets of BIAcore binding curves.

Figure 2A shows an overlay of binding curves (sensograms) obtained after passing Fab TVG405 over a BIAcore chip coated with MBP-E4 fusion protein as described above. Fab concentrations range from 10mM (lowest curve) to 300nM (upper curve) through 5 intermediate dilutions. The extent of binding is indicated in resonance units on the X-axis, against time in seconds on the Y-axis. Purified Fab is injected at around 100

seconds and washing initiated at 700 seconds. The affinity (K_d) of TVG405 is calculated as between 0.3 and 1.25nM by analysis of the association and dissociation curves using BIAevaluation software (Pharmacia, UK).

Figure 2B shows an overlay of binding curves (as described above) for the hybridoma-derived Fab TVG402 over a concentration range 100nM to 1 M. The K_d is estimated as 70nM.

TVG405 has an association rate constant (k_{on}) of 1.8 x 10^6 $M^{-1}.s^{-1}$ and an off rate (k_{off}) of 2 x 10^3 s^{-1} indicating a molar dissociation constant (Kd) of approximately 1nM. The best hybridoma-derived antibody - TVG402 - has an affinity of only 70nM, and had a k_{on} of 4.2x10^4 $M^{-1}.s^{-1}$ and a k_{off} value of 3x10^3 s^{-1}. TVG 406 and 407 display rapid kinetics and are thus examined by Scatchard analysis of equilibrium binding data, as shown for TVG407.

Figure 2C shows the equilibrium binding curve of Fab TVG407, which displays rapid kinetics. Figure 2D shows Scatchard analysis of the data presented in Fig. 2C using BIAevaluation software. Equilibrium values are corrected for bulk refractive index changes by subtracting values from a surface blocked with biotin, from the values shown in Fig. 2C. In the plot shown the slope is - K_d and the Y-axis intercept is 'R_{max}', i.e. the binding level at saturation with Fab. The uncorrected K_d values for TVG407 and TVG406 are 250nM and 140nM which, when the activity of the Fab preparation is considered, indicates actual affinities of 50nM and 28nM.

TVG407 has an affinity (Kd) of 50nm after correction for biological activity, and TVG406 has an affinity (Kd) of 28nM. The amino acid sequence of the heavy and light chain CDR3 loops are established by DNA sequencing, further confirming that the three antibodies are distinct.

Example 3

Preparation of Anti-E4 Peptides

A commercially available two-hybrid screening kit is purchased from ClonTech and employed for identifying naturally occurring E4-binding peptides, according to the instructions given by the manufacturer. A HeLa cDNA library, obtained from the same supplier, is screened. By this method, seven DNA sequences are isolated which encode E4-binding polypeptides, of which three are identified after sequencing.

The first peptide is ferritin.

The second peptide is a keratin filament binding protein, which has the sequence set forth in SEQ. ID. No. 2.

The third polypeptide is a novel polypeptide recognised as a member of the DEADbox family of proteins, which contain the characteristic sequence motif DEAD. The sequence of the third polypeptide is shown in SEQ. ID. No. 3.

In order to identify the site of interaction between these polypeptides and E4, a series of overlapping peptides of between 10 and 20 amino acids in length is generated by PCR and displayed on phage as described above. The binders are subsequently employed as screening agents to identify HPV16 in mucosal lesions.

Example 4

Preparation of Anti-E4 RNA oligonucleotides

RNA oligonucleotides, known as aptamers, which are capable of specific binding to target molecules can be generated by selection procedures such as SELEX. The SELEX method involves selection of nucleic acid aptamers, single-stranded nucleic acids capable of binding to a desired target, from a library of oligonucleotides. Starting from a library of nucleic acids, preferably comprising a segment of randomised sequence, the SELEX method includes steps of contacting the library with the target under conditions favourable for binding, partitioning unbound nucleic acids from those

nucleic acids which have bound specifically to target molecules, dissociating the nucleic acid-target complexes, amplifying the nucleic acids dissociated from the nucleic acid-target complexes to yield a ligand-enriched library of nucleic acids, then reiterating the steps of binding, partitioning, dissociating and amplifying through as many cycles as desired to yield highly specific, high affinity nucleic acid ligands to the target molecule.

DNA Oligonucleotide Library

DNA oligonucleotides 73 bases in length, having a randomised portion of 26 bases. are used for the development of an aptamer capable of binding E4. A library of synthetic RNA oligonucleotides having the following structure is prepared:

5' CCTGTTGTGAGCCTCCTGTCGAA(26N)TTGAGCGTTTATTCTTGTCTCCC 3'

Where N stands for any possible base in the random region. The random region is generated by using a mixture of all four nucleotides (ratio 6:5:5:4, A:C:G:T, to allow for differences in coupling efficiency) during the synthesis of each nucleotide in that stretch of the oligonucleotide library. The resulting complexity is theoretically 4^{26} molecules. The scale of synthesis (0.1µmol) followed by gel purification yields 8.8nmol which puts an absolute upper limit of approximately $5x10^{15}$ on the number of ...ıt molecules actually present.

...mplification with ...the recognition site for T7 RNA ...-ase (5' TAATA... ...CAAGAATAAACGCTCAA 3') ...rimer (5' GCCTC... ...AA 3') results in the following for transcription:

5' TAATAGCACTCACTATAGGGAGACAAGAATAAACGCTCAA (26N)

TTCGACAGGAGGCTCACAACAGGC 3'

The RNA transcript itself has the following sequence:

5' GGGAGACAAGAAUAAACGCUCAA (26N) UUCGACAGGAGGCUCACAACAGGC 3'

Anti-E4 aptamers are selected using a conventional SELEX procedure as described in US Patent 5,270,163. Each round consists of the following steps:

1) Selection. The RNA and E4 protein are mixed, incubated at 37° C., washed through a nitrocellulose filter, and RNA is eluted from the filters.

2) Amplification. The RNA eluted from filters is extended with AMV reverse transcriptase in the presence of 50 picomoles of 3' primer in a 50 µl reaction under conditions described in Gauss et al. (1987). To the resulting cDNA synthesis 50 picomoles of 5' primer is added and in a reaction volume of 100µl and amplified with Taq DNA polymerase as described in Innis (1988) for 30 cycles.

3) Transcription. In vitro transcription is performed on the selected amplified templates as described in Milligan et al. (1987), after which DNaseI is added to remove the DNA template. The resultant selected RNA transcripts are then used in step 1 of the next round. Only one-twentieth of the products created at each step of the cycle are used in the subsequent cycles so that the history of the selection can be traced. The progress of the selection method is monitored by filter binding assays of labeled transcripts from each PCR reaction. After the fourth round of selection and amplification, the labeled selected RNA products produce binding to E4. The binders are used in the detection of HPV in cells derived from cervical smears.

Example 5

Detection of HPV in Cutaneous and Mucosal Lesions

All the synthetic Fabs detect the HPV16 E1^E4 protein in formalin fixed paraffin-embedded tissue, although TVG405 consistently show the highest level of staining (Figure 3).

Figure 3 illustrates the use of synthetic Fabs to localise HPV16 E4 protein *in vivo* by immunostaining of a low grade HPV16 CIN I with Fab NIP-C11 (Griffiths *et al*, 1994), which has no reactivity towards HPV16 E4 (Fig. 3A), and the E4-specific Fab TVG405 which is described here (Figs. 3B, C, D). Fabs are detected using 9E10 as secondary antibody followed by anti-mouse FITC conjugate. E4 is detectable in the upper layers of the epidermis but at greatly varying levels between different lesions with often only a few positive cells being apparent (C, D). The position of the basal layer is arrowed in C and D. Magnification is 200X.

Epitope exposure by microwave treatment enhances the sensitivity of E4 detection, and even allows staining using TVG402 (Doorbar *et al*, 1992). The extent of E4 expression varies greatly between different lesions (8 HPV16-associated CIN1 biopsies are examined), ranging from expression only in rare cells scattered throughout the biopsy (Fig. 3), to widespread distribution throughout the most differentiated layers of the epidermis (Fig. 4), comparable to the distribution of E4 in cutaneous warts caused by HPV1 and HPV63 where the production of infections virions is also high (Fig. 4). In low grade cervical intraepithelial neoplasia (CIN 1) caused by HPV16, E4 and L1 levels are also found to correlate closely, although expression of the two proteins is not coincident (as previously suggested (Brown et al, 1994). E4 expression precedes the synthesis of the major capsid protein by several cell layers (as revealed by double staining, see Fig. 4) and in high grade cervical lesions (CIN 2/CIN 3) E4 is often abundant, even though the expression of L1 is no longer supported (Fig 4). This time delay between the commencement of E4 synthesis and the assembly of infectious virions is most apparent in HPV63, where E4 expression coincided with migration of an infected basal cell into the parabasal layers, while expression of L1 is restricted to a narrow strip of cells in the upper granular layer.

Figure 4 demonstrates that synthesis of E4 is not directly linked to the expression of capsid proteins in high and low grade HPV16 lesions, and benign warts. Figure 4 shows the results of triple staining using anti L1 antisera (Figs. 4A, D, G), HPV16 E4

Fab TVG405 (Figs. 4B and 4E), polyclonal antisera to HPV63 E4 (Fig. 4H), and with DAPI (Figs. 4C, F, I). A, B and C represent a low grade HPV16-induced lesion (CIN I). D, E and F represent a high grade HPV16-induced lesion (CIN II/III). G, H and I represent a section through a verruca caused by HPV63. In all cases E4 expression precedes L1 expression although by only a few cell layers in CIN I (A, B). In the CIN II/III we assume that terminal differentiation is insufficient to support synthesis of virion structural proteins (D) although E4 expression is abundant (E). The contrast between the onset of E4 expression and the detection of virus structural proteins is most apparent in cutaneous verrucas caused by HPV63 (G, H). The basal layer is indicated by an arrow on the DAPI-stained images. Magnification is 100X.

Onset of vegetative viral DNA replication and expression of E4 coincide

Vegetative viral DNA replication is found to begin in cells of the mid spinous layer and to correlate exactly with the onset of E4 expression (Fig 5).

Figure 5 demonstrates that onset of vegetative viral DNA replication coincides with E4 expression in low grade HPV16 lesions and in benign cutaneous warts. The figure shows triple staining using the HPV16 E4 antibodies TVG402, 405 and 406 (Fig. 5A) and HPV1 E4 antibodies 4.37 and 9.95 (Fig. 5D), biotinylated DNA probe (Fig. 5B - HPV16, Fig. 5E - HPV1), or DAPI (Figs. 5C and F). A, B and C represent a section through an HPV16-induced CIN I, and D, E and F represent a section through an HPV1-induced verruca. In the HPV16 CIN I, vegetative viral DNA replication and E4 synthesis correlate in the mid to upper layers of the epidermis (A, B). In cutaneous lesions the two events are initiated as soon as the infected cell leaves the basal layer (D, E). Basal cells are illustrated in the DAPI counterstained image (F). Magnification is 200X.

In HPV 1-induced warts vegetative viral DNA replication and E4 synthesis commence much earlier, and are evident immediately after the infected basal cell migrates into the superficial layers (Fig 5). Only a proportion of the differentiating cells are permissive for vegetative viral DNA replication, and only in these cells is E4 detectable.

Neighbouring cells showed neither late gene expression nor vegetative viral DNA replication, suggesting that onset of the two events is closely linked. Although the sensitivity of DNA and E4 detection is not established, these 'normal' cells are likely to be either non-permissive for viral replication or be uninfected. This precise correlation between E4 expression and the onset of vegetative viral DNA replication is also seen in cutaneous warts caused by HPV63 and 65, and in common warts caused by HPV2.

Cells undergoing late gene expression show an abnormal pattern of terminal differentiation when compared to non-permissive or uninfected cells

Cells supporting the late stages of HPV infection can thus be identified by immunostaining with Fab TVG405 (for HPV16) Mab 4.37 (for HPV1) or polyclonal antisera to E4 (HPV63). In warts caused by HPV1, E4-positive cells lack detectable levels of filaggrin or involucrin (Fig 6(i)). Non-permissive (or uninfected) cells in the same lesion which show neither E4 expression nor vegetative viral DNA replication, express filaggrin and loricrin at levels indistinguishable from those in the surrounding epidermis. Correlation of E4 synthesis with the differentiation-specific keratins K4 and K13 reveals an identical pattern of inhibition. The intensity of K4 and K13 staining is always lower in E4-positive cells than in neighbouring cells that are not expressing E4 (Fig.6(ii)). K5 and 14, which are present in the basal and lower parabasal cells, are unaffected. This interference with the detection of expression differentiation-specific keratins (K1 and K10 in cutaneous skin) is also apparent in cutaneous warts caused by HPVI (Fig 6(ii)) but is not evident in warts caused by HPV63 (Fig 6(ii)). The E4 protein of HPV63 is most closely related to that of HPV1.

Figure 6 illustrates that productive infection interferes with normal epithelial terminal differentiation in low grade HPV16 lesions and in benign cutaneous warts. Figure 6(i) (keratin expression) shows triple staining using the HPV16 E4 Fabs TVG405/TVG406 (Fig. 6(i)A), HPV 1 E4 monoclonals 4.37/9.95 (D), and HPV63 E4 polyclonal antibodies (G), in conjunction with antibodies to the differentiation-specific mucosal keratins 4 and 13 (B) or cutaneous keratins 1 and 10 (E, H). Figures 6(i) C, F and I show the DAPI counter stain. A, B and C represent a section through a HPV16-

induced CIN I. D, E and F show a section through the edge of an HPV1-induced verruca, while Figures 6(i) G, H and I show a section through an HPV63-induced wart. In HPV16 and HPV1-induced lesions, differentiation-specific keratins are less apparent in E4-positive cells than in neighbouring cells (A, B, D, E) although this is not the case with HPV63 (G, H). Nuclear degeneration (visualised by DAPI counter staining) is retarded in E4-expressing cells (A, C, D, F). Magnification is 200X.

Figure 6(ii) relates to filaggrin expression. The figure shows triple staining, as described above, except that Figures 6(ii) B and E show filaggrin staining. E4 staining is shown in figures 6(ii) A and D, and DAPI counter staining is shown in figures 6(ii) C and F. A, B and C represent the edge of an HPV63-induced wart where normal skin (left hand side of figure) meets the benign tumour (right hand side of figure). D, E and F show the granular layer of an HPV1-induced wart. E4-positive cells do not express detectable levels of the differentiation-specific marker filaggrin, and show marked nuclear preservation when compared to neighbouring uninfected or non-permissive cells. Magnification is 200X.

The intracellular distribution of the HPV16 E4 proteins is distinct from the distribution of E4 in cutaneous lesions caused by HPV1 and HPV63.

The E1^E4 protein of HPV1 is predominantly cytoplasmic and assembles into inclusions that coalesce and increase in size as the cell migrates towards the surface of the skin. The E1^E4 protein of HPV63 is found to have a fibrous and granular distribution. By contrast, HPV16 E4 had a filamentous and perinuclear distribution in cells of the lower epidermal layers (Fig, 7), and assembled into prominent structures only in the more differentiated cell layers. These 'inclusions' are always found singly per cell (c.f. multiple inclusions found in most cutaneous lesions), are located adjacent to the nucleus, and are nearly always detected on the side of the nucleus closest to the surface of the epidermis. Although similar in appearance to the E4/intermediate filament bundles which form after expression of the HPV16 E1^E4 protein in epithelial cells *in vitro*, we have not detected the presence of keratins in these structures *in vivo*. Antibodies to the very N-terminus of HPV16 E1^E4 stained the structures much less

readily than antibodies to C-terminal epitopes (TVG 404, TVG405, TVG406) suggesting that the N-terminal region maybe either hidden or lost.

Figure 7 shows the association of the HPV16 E4 proteins with perinuclear bundles and filamentous structure *in vivo*, in particular the detection of HPV16 E4 proteins in the upper layers (Figs. 7A, B) and lower layers (Figs. 7C, D) of a HPV16 CIN I using a mixture of Fabs TVG405 and TVG406. In the upper layers E4 is diffuse throughout the cytoplasm but with a prominent perinuclear pattern. Concentration of E4 into perinuclear bundles (arrowed in Fig. 7B) is apparent in these cells. In the lower layers, E4 has a predominantly perinuclear and filamentous appearance (Figs. 7C, D), but is not concentrated into perinuclear bundles. Magnification for Figs. 7A and C is 200X; that for B and D is 400X.

Confocal imaging revealed the N-terminal antibodies to localise primarily to the edge of the E4 structures while anti C-terminal staining is strongest in the centre (data not shown). When compared to the distribution seen with TVG405 and TVG406, the anti N-terminal reagent revealed HPV16 E1^E4 to have a more diffuse distribution in the cell (Fig. 8). No significant difference is apparent between the staining pattern of TVG405, 406, 407 and the C-terminal antibody.

Figure 8 provides evidence for processing of the HPV16 E4 proteins *in vivo* and shows triple staining in the upper layers of a HPV16 CIN using HPV16 E4 Fab TVG406 which recognises an epitope in the C-terminal half of the E4 protein (Fig. 8A), an antibody to the N-terminal 12 amino acids of the HPV16 E1^E4 protein (Fig. 8B) and DAPI (Fig. 8C). TVG402, 403, 404, 405 and 407 produced staining patterns that are not significantly different from that of TVG 406. Anti N-terminal antibodies did not effectively stain the perinuclear bundles (8B) which are evident with TVG406 (arrowed in 8A) suggesting that as in HPV1, different forms of the protein have different intracellular locations. Magnification is 400X.

Example 6

Detection of HSV in cells isolated from cervical lesions

Slides suitable for imaging of cells derived from cervical smears stained using anti-E4 antibodies are prepared by the method set forth in US 5,346,831. Cells are isolated from a patient according to conventional procedures and dissolved in 10ml alcohol/saline buffer. The sample is prepared for centrifugation by disaggregating the clumps or clusters of cells in the sample vial by vortexing. After disaggregation, the sample is drained completely and layered over a density gradient in a 12 ml conical tube, wherein the density gradient is formed with a plasma expander material comprising 6% betastarch solution, and 0.9% physiological saline, also known by the tradename "Hespan" made by NPBI, Emmer-Compascuum, the Netherlands.

12 ml conical tubes containing density gradient and sample cells are placed into centrifuge buckets, balance and centrifuged for 5 minutes, at a force of about 600G. The liquid is then aspirated down to the 5 ml mark on the conical tube. The centrifuge buckets are removed and the 12 ml conical tube centrifuged with remaining liquid for 10 minutes, at 800G. The tubes are emptied of supernatant, tapping lightly 2 or 3 times at a 45 degree angle. The tubes now contain packed cells of varying volumes. Upon mixing to homogeneity, the pellets generally contain the same concentration of cells per unit volume of liquid.

50µl of deionized H_2O is added, and the sample mixed by syringing 5 times through a 0.042 inch tip. Upon completion of mixing, 150 µl of sample followed by 500 µl of deionized H_2O is dispensed into a sedimentation vessel attached to a slide which has been conventionally coated with Poly-L lysine (1% Sigma). The transferred sample is allowed to settle within the vessel for approximately 10 minutes. The excess sample is aspirated off and the chamber rinsed with 2 ml deionized H_2O two times (aspirating between each addition).

FITC-labelled Fabs are then applied to the cells according to known procedures and the binding visualised by fluorescence microscopy.

References

Andrews & DiMaio, (1993) *J. Virol.* **67**, 7705-7710.

Barksdale & Baker (1995). *J. Virol* **69**, 6553-6556.

Beyer-Finkler *et al*, (1990) *Med. Microbiol. Immunol.* **179**, 185-192.

Breitburd *et al*, (1987) *In* "Cancer Cells 5", pp. 115-122. cold Spring Harbor Laboratory press. Cold Spring Harbor, New York.

Brown *et al*, (1994) *Virology* **201**, 46-54.

Chow & Broker (1994), *Intervirology* **37**, 150-158.

Chow *et al*, (1987) *J. Virol.* **61**, 2581-2588.

Croissant *et al*, (1985) *Clin. Dermatol.* **3**(4), 43-55.

Crum *et al*, (1990) Virology **178**, 238-246.

Dietrich-Goetz *et al*, (1997) *Proc. Natl. Acad. Sci. USA* **94**, 163-168.

Doorbar, J. (1996). The E4 proteins and their role in the viral life cycle. *In* "Papillomavirus Reviews: Current Research on Papillomaviruses" (C. Lacey, Ed.), pp. 31-38. Leeds Medical Information, Leeds University Press, Leeds.

Doorbar *et al*, (1989). *EMBO J.* **5**(2), 355-362.

Doorbar *et al*, (1989). *Virology* **172**, 51-62.

Doorbar *et al*, (1992). **187**, 353-359.

Doorbar *et al*, (1991) *Nature* **352**, 824-827.

Doorbar *et al*, (1988). *EMBO J.* 73, 825-833.

Doorbar & Gallimore (1987). *J. Virol* **61**, 2793-2977.

Doorbar *et al*, (1996a). *Virology* **218**, 114-126.

Doorbar & Myers (1996b). The E4-protein. *In* "Human Papillomaviruses 1996" (G. Myers, H. Delius, J. Icenogel, H.-U. Bernard, C. Baker, A. Halpern, and C. Wheeler, Eds), Vol. III, pp. 58-80. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

Egawa (1994) *Brit. J. Dermatol.* **130**, 158-166.

Egawa *et al*, (1993a). *Virology* **194**, 51-62.

Egawa *et al*, (1993b). *Brit. J. Dermatol.* **128**, 271-276.

Frattini *et al*, (1996). *Proc. Natl. Acad. Sci. USA.* **93**(7), 3062-3067.

Furth & Baker (1991). *J. Virol.* **65**, 5806-5812.

Furth *et al*, (1994). *Mol. Cell Biol.* **14**(8), 5278-5289.

Grand *et al*, (1989). *Virology* **170**, 201-213.

Green *et al*, (1982). *Cell* **28**, 477-487.

Griffiths *et al*, (1994). *EMBO J.* **13**, 3245-3260.

Hawkins *et al*, (1992). *J. Mol. Biol.* **226**, 889-896.

Hudson *et al*, (1992). *Hybridoma* **11**(3), 367-378.

Hummel *et al*, (1995). *J. Virol.* **69**, 3381-3388.

Jareborg & Burnett (1991). *J. Gen. Virol.* **72**, 2269-2274.

Johnson *et al*, (1991). *Analyt. Biochem.* **198**, 268-277.

Kennedy *et al*, (1991). *J. Virol.* **65**, 2093-2097.

Laimins (1993). The Biology of Human Papillomaviruses: From Warts to Cancer. *Infectious Agents and Disease* **2**, 74-86.

Lambert (1991). *J. Virol.* **65**, 3417-3420.

Low *et al*, (1996). *J. Mol. Biol.* **260**(3), 359-368.

Marks *et al*, (1992). *Bio-technology* **10**(7), 779-783.

McClean *et al*, (1990) *J. Clin. Pathol.* **43**, 488-492.

Meyers *et al*, (1992). *Science* **257**, 971-973.

Nissim *et al*, (1994). *EMBO J.* **13**, 692-698.

Palefsky *et al*, (1991). *J. Clin. Invest.* **87**, 2132-2141.

Pope *et al*, (1996). *Immunotechnology* **2**(3), 209-217.

Pray *et al*, (1995). *Virology* **206**, 679-685.

Roberts & Wientraub (1996). *Cell* **46**, 741-752.

Roberts *et al*, (1994). *J. Virol* **68**(10), 6432-6455.

Roberts *et al*, (1993). *Virology* **197**, 176-187.

Rogel-Gaillard *et al*, (1992). *J. Virol.* **66**(2), 816-823.

Rogel-Gaillard *et al*, (1993). *J. Invest. Dermatol.* **101**, 843-851.

Schier *et al*, (1996). *J. Mol, Biol.* **263**(4), 551-567.

Schneider (1994). *Intervirology* **37**(3-4), 201-214.

Sherman & Schlegel (1996). *J Virol.* **70**, 3269-3279.

Stoler *et al*, (1990). *J. Virol.* **64**, 3310-3318.

Stoppler et al, (1996). *J. Virol.* **70**,6987-6993.

Villiers de (1994) Human pathogenic papillomavirus types: an update. *In* "Human Pathogenic Papillomaviruses" (H. Zur Hausen, Ed). ppl 1-12. Springer-Verlag, New York.

Zheng *et al*, (1996). *J. Virol.* **70**, 4691-4699.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: MEDICAL RESEARCH COUNCIL
(B) STREET: 20 PARK CRESCENT
(C) CITY: LONDON
(E) COUNTRY: GB
(F) POSTAL CODE (ZIP): W1N 4AL

(ii) TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA VIRUSES

(iii) NUMBER OF SEQUENCES: 3

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 375 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GCG CTG CCA CTC TCA GAA GTT ATT GTC AAA AAC TTG CAA CTT GCT TTG        48
Ala Leu Pro Leu Ser Glu Val Ile Val Lys Asn Leu Gln Leu Ala Leu
 1               5                  10                  15

GCA AAT AGC TCT CGA AAT GCT GTC GCT CTT TCT GCC AGC CCT CAA CTG        96
Ala Asn Ser Ser Arg Asn Ala Val Ala Leu Ser Ala Ser Pro Gln Leu
            20                  25                  30
```

```
AAA GAG GCC CAG TCA GAG AAG GAA GAA GCC CCA AAG CCA CTT CAC AAA        144
Lys Glu Ala Gln Ser Glu Lys Glu Glu Ala Pro Lys Pro Leu His Lys
          35                  40                  45

GTA GTG GTA TGT GTT AGT AAA AAA CTC AGT AAG AAG CAG AGT GAA CTA        192
Val Val Val Cys Val Ser Lys Lys Leu Ser Lys Lys Gln Ser Glu Leu
     50                  55                  60

AAT GGG ATC GCA GCC TCT CTA GGA GCA GAT TAC AGG TGG AGT TTT GAT        240
Asn Gly Ile Ala Ala Ser Leu Gly Ala Asp Tyr Arg Trp Ser Phe Asp
 65                  70                  75                  80

GAA ACA GTG ACT CAT TTC ATC TAT CAA GGG CGG CCA AAT GAC ACT AAT        288
Glu Thr Val Thr His Phe Ile Tyr Gln Gly Arg Pro Asn Asp Thr Asn
                 85                  90                  95

CGG GAG TAT AAA TCT GTA AAA GAA AGA GGA GTA CAC ATT GTT TCC GAG        336
Arg Glu Tyr Lys Ser Val Lys Glu Arg Gly Val His Ile Val Ser Glu
            100                 105                 110

CAC TGG CTT TTA GAT TGT GCC CAA GAG TGT AAA CAT CTT                    375
His Trp Leu Leu Asp Cys Ala Gln Glu Cys Lys His Leu
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Ala Leu Pro Leu Ser Glu Val Ile Val Lys Asn Leu Gln Leu Ala Leu
  1               5                  10                  15

Ala Asn Ser Ser Arg Asn Ala Val Ala Leu Ser Ala Ser Pro Gln Leu
            20                  25                  30

Lys Glu Ala Gln Ser Glu Lys Glu Glu Ala Pro Lys Pro Leu His Lys
          35                  40                  45

Val Val Val Cys Val Ser Lys Lys Leu Ser Lys Lys Gln Ser Glu Leu
     50                  55                  60

Asn Gly Ile Ala Ala Ser Leu Gly Ala Asp Tyr Arg Trp Ser Phe Asp
 65                  70                  75                  80

Glu Thr Val Thr His Phe Ile Tyr Gln Gly Arg Pro Asn Asp Thr Asn
                 85                  90                  95

Arg Glu Tyr Lys Ser Val Lys Glu Arg Gly Val His Ile Val Ser Glu
            100                 105                 110
```

His Trp Leu Leu Asp Cys Ala Gln Glu Cys Lys His Leu
115 120 125

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 491 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Ala Pro Glu Glu His Asp Ser Pro Thr Glu Ala Ser Gln Pro Ile Val
1 5 10 15

Glu Glu Glu Glu Thr Lys Thr Phe Lys Asp Leu Gly Val Thr Asp Val
20 25 30

Leu Cys Glu Ala Cys Asp Gln Leu Gly Trp Thr Lys Pro Thr Lys Ile
35 40 45

Gln Ile Glu Ala Tyr Ser Leu Ala Leu Gln Gly Arg Asp Ile Ile Gly
50 55 60

Leu Ala Glu Thr Gly Ser Gly Lys Thr Gly Ala Phe Ala Leu Pro Ile
65 70 75 80

Leu Asn Ala Leu Leu Glu Thr Pro Gln Arg Leu Phe Ala Leu Val Leu
85 90 95

Thr Pro Thr Arg Ser Trp Pro Phe Arg Ser Gln Ser Ser Leu Lys Pro
100 105 110

Trp Ser Ser Ile Gly Val Gln Ser Ala Val Ile Val Gly Gly Ile Asp
115 120 125

Ser Met Ser Gln Ser Leu Ala Leu Ala Lys Lys Pro His Ile Ile Ile
130 135 140

Ala Thr Pro Gly Arg Leu Ile Asp His Leu Glu Asn Thr Lys Gly Phe
145 150 155 160

Asn Leu Arg Ala Leu Lys Tyr Leu Val Met Asp Glu Ala Asp Arg Ile
165 170 175

Leu Asn Met Asp Phe Glu Thr Glu Val Asp Lys Ile Leu Lys Val Ile
180 185 190

Pro Arg Asp Arg Lys Thr Phe Leu Phe Ser Ala Thr Met Thr Lys Lys
195 200 205

Val Gln Lys Leu Gln Arg Ala Ala Leu Lys Asn Pro Val Lys Cys Ala
210 215 220

Val Ser Ser Lys Tyr Gln Thr Val Glu Lys Leu Gln Gln Tyr Tyr Ile
225 230 235 240

Phe Ile Pro Ser Lys Phe Lys Asp Thr Tyr Leu Val Tyr Ile Leu Asn
245 250 255

Glu Leu Ala Gly Asn Ser Phe Met Ile Phe Cys Ser Thr Cys Asn Asn
260 265 270

Thr Gln Arg Thr Ala Leu Leu Leu Arg Asn Leu Gly Phe Thr Ala Ile
275 280 285

Pro Leu His Gly Gln Met Ser Lys Arg Leu Gly Ser Leu Asn Lys Phe
290 295 300

Lys Ala Lys Ala Arg Ser Ile Leu Leu Ala Thr Asp Val Ala Ser Arg
305 310 315 320

Gly Leu Asp Ile Pro His Val Asp Val Val Val Asn Phe Asp Ile Pro
325 330 335

Thr His Ser Lys Asp Tyr Ile His Arg Val Gly Arg Thr Ala Arg Ala
340 345 350

Gly Arg Ser Gly Lys Ala Ile Thr Phe Val Thr Gln Tyr Asp Val Glu
355 360 365

Leu Phe Gln Arg Ile Glu His Leu Ile Gly Lys Lys Leu Pro Gly Phe
370 375 380

Pro Thr Gln Asp Asp Glu Val Met Met Leu Thr Glu Arg Val Ala Glu
385 390 395 400

Ala Gln Arg Phe Ala Arg Met Glu Leu Arg Glu His Gly Glu Lys Lys
405 410 415

Lys Arg Ser Arg Glu Asp Ala Gly Asp Asn Asp Asp Thr Arg Gly Cys
420 425 430

Tyr Val Cys Gln Glu Gln Gly Gly Trp Arg Lys Asn Glu Glu Ala Glu
435 440 445

Arg Pro Leu Ile Thr Phe Met Lys Ala Arg Val Leu Leu Phe Cys Lys
450 455 460

Arg Glu Leu Glu Asn Glu Thr Cys Ser Asn Arg Asp His Glu Thr Glu
465 470 475 480

Ile Gly Gln Asn Cys Val Gln Asn Val Leu Ser
485 490